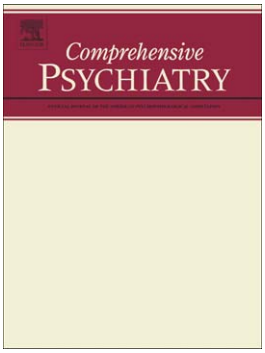

The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients

Laura E. O'Shea, Geoffrey L. Dickens

PII: S0010-440X(15)00122-4
DOI: doi: 10.1016/j.comppsych.2015.07.009
Reference: YCOMP 51551

To appear in: *Comprehensive Psychiatry*

Received date: 20 February 2015
Revised date: 6 July 2015
Accepted date: 14 July 2015

Please cite this article as: O'Shea Laura E., Dickens Geoffrey L., The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients, *Comprehensive Psychiatry* (2015), doi: 10.1016/j.comppsych.2015.07.009

The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients

Laura E O'Shea[ab], Geoffrey L Dickens*[ac]

[a]St Andrew's Academic Department, Billing Road, Northampton, NN1 5DG, United Kingdom

[b]Institute of Psychiatry, Psychology & Neuroscience, King's College London, 16 De Crespigny Park, London, SE5 8AF, United Kingdom

[c]School of Social and Health Sciences, Abertay University, Bell Street, Dundee, DD1 1HG, United Kingdom

***Correspondence:** Geoffrey L Dickens, School of Health and Social Sciences, Abertay University, Bell Street, Dundee, DD1 1HG, United Kingdom, Email g.dickens@abertay.ac.uk. Tel: +441382 308257

Abstract

Objective: Periodic structured violence risk assessment is the principle method underlying treatment planning for mentally disordered offenders but little is known about how risk changes over time. We aimed to determine whether hospitalised patients underwent reliable clinical change in assessed risk.

Method: We used a pseudo-prospective longitudinal study design. Demographic, clinical and risk assessment data of adult inpatients (*N*=480) who had been routinely assessed with the HCR-20 on two to four occasions over a mean period of 17 months (SD=2) were collated. Linear mixed models regression was conducted to determine change over time on total, subscale, and individual item scores, and relative change between clinical and demographic groups. The Reliable Change Index was calculated to examine whether change was greater than that expected by measurement error; clinically significant change was defined as the extent to which HCR-20 scores reduced below previously reported scores for patients not requiring hospitalisation.

Results: HCR-20 total score (Estimate -0.42, 95% CI=-0.84, -0.01, $p<.05$; $d=.20$) and clinical score (Estimate=-0.42, 95% CI=-0.64, -0.20, $p<.001$; $d=.36$) reduced over assessments. Significant differences in change were evident between clinically and demographically defined groups. A maximum of 3% of individuals showed clinically significant reliable reductions in HCR-20 total scores. The scores of patients whose overall level of risk was judged to have decreased did not reduce between assessments.

Conclusion: Violence risk changes very little over the course of treatment although there is some variation between groups. Most change cannot be demonstrated to be reliable or clinically significant. Important clinical management decisions should not depend solely on evidence from changes in HCR-20 risk assessment.

Keywords: Risk assessment, violence, reliable change index, schizophrenia, personality disorder, intellectual disability

1. Introduction

[1]Structured professional judgment is the gold standard approach to violence risk assessment (1) and bridges the gap between unstructured clinical opinion and the purely actuarial. Whilst historical factors should form the basis of all risk assessments (2), the variable or dynamic risk factors that both precede violence and are potentially modifiable (3) are the most promising targets for interventions aimed at reducing future violence (4) and have therefore become a popular aspect of risk assessment schemes. Version 2 of the Historical, Clinical, Risk Management-20 (HCR-20; 5) is the most commonly used structured professional judgment tool in secure psychiatric units in the United Kingdom (6). It comprises 10 historical factors that are thought to be relatively static insomuch as once they are judged to be present they will remain so in most circumstances unless new behaviour is exhibited or new information comes to light (7); and 10 dynamic factors, which are divided into a clinical scale containing items about current clinical presentation, and a risk management scale comprising items about adjustment to future risk-related circumstances; each item is scored as present (2), possibly present (1), or absent (0) resulting in a maximum score of 40. The majority of the research into the HCR-20 has examined its predictive validity; such studies have been summarised in an annotated bibliography maintained by the tools authors (8). Studies of its predictive ability specifically among psychiatric inpatients have been synthesised in a recent meta-analysis (9).

The ultimate aim of risk assessment is to identify management strategies and interventions to reduce identified risk factors and ultimately reduce the incidence of violence (4). It is therefore surprising that few studies have examined the ability of the HCR-20 to detect change in risk factors despite recommendations that the HCR-20 be repeated every 6 to

[1] HCR-20, Historical, Clinical, Risk-management-20; OAS, Overt Aggression scale; CI, confidence interval; DSPD, dangerous and severe personality disorder; RCI, reliable change index; R&CSC, reliable and clinically significant change; DUNDRUM, dangerousness, understanding, recovery and urgency manual

12 months (5). Those studies that have examined change in scores over time have been conducted in high secure settings. Longer treatment times have been found to be associated with reduced scores on the clinical and risk-management scales (10, 11) and mean scores reduce between the first and second HCR-20 assessments (12). Morrissey and colleagues (13) also found statistically significant reductions in dynamic items each year. However, across studies, average changes were small with total clinical scores reducing by 0.19 to 0.32 points and risk-management scores reducing by 0.14 to 0.31 points per iteration (12, 13). Examination of change in individual items by Olsson et al. (12) revealed that all individual clinical items reduced significantly, with the exception of C4 (impulsivity). Of the risk management items, R1 (plans lack feasibility) and R4 (noncompliance with remediation attempts) also reduced significantly. Current evidence suggests that women's scores reduce more than men's over time. Olsson et al. (12) found that women's scores were 2.2 times more likely than men's to reduce by 30% or more between assessments, and Morrissey et al. (13) found that scores for women with a range of diagnoses decreased by an additional 0.23 points per year on the clinical scale relative to patients in a male mental illness service. In contrast, risk-management scores for the male patients resident in the Dangerous and Severe Personality Disorder (DSPD) and intellectual disability services increased relative to the mental illness group (13).

1.1. Contribution of the current study

Previous research has demonstrated that clinical and risk-management scores reduce over time; however, the majority of the research has been conducted in high secure psychiatric settings where individuals will likely have higher baseline scores on risk assessments. Further, actual changes in scores have been modest and may not represent reliable or clinically significant change. Therefore, the current study aims to extend previous findings by investigating changes in HCR-20 ratings over time among patients resident in

medium and low secure services and determining whether any observed reduction in scores are reliable (i.e. unlikely to be due to measurement error) and clinically significant (i.e. scores reduce from those representative of a dysfunctional population to those of a relatively normal population). Previous evidence suggests that there may be differences in rates of change between services, or demographically and clinically defined groups; specifically, that women show a greater reduction in scores compared with men. Therefore, the current study aimed to further investigate changes in dynamic risk among patients in secure psychiatric services by examining which sample characteristics affect rate of change in HCR-20 scores. We also aimed to examine differences in rates of change of individual HCR-20 items between groups as the items that are most amenable to change for a given group may be most usefully prioritised as treatment targets.

2. Method

2.1. Setting

St. Andrew's is a charitable provider of inpatient secure mental health care in England. Gender-specific, medium and low secure accommodation is provided in specialist diagnostic or demographically defined care pathways (e.g. older adults, autistic spectrum disorder).

2.2. Participants

To be considered eligible, patients had to be aged 18 years or over, resident in any service between September 2010 and March 2011, and have at least two HCR-20 assessments completed by their clinical team. HCR-20 assessments were excluded if they were missing data exceeding prorating guidelines in the HCR-20 manual (5).

2.3. Design and procedure

The study was approved by the clinical audit and service evaluation committee and employed a pseudo-prospective design; HCR-20 assessments were completed as part of routine clinical practice by registered psychologists or graduate psychology assistants under their supervision prior to ratification by the clinical team. Risk assessment data, corresponding to up to four HCR-20 assessments per patient, and demographic and clinical information were extracted from anonymised patient records.

2.4. Measures

2.4.1. HCR-20

Risk assessment was completed with version 2 of the HCR-20 (5), as version 3 (14) was not implemented in the study setting at the time of data collection. The HCR-20 comprises 20 items; 10 historical scale items that are thought to be relatively static, five clinical items that reflect the individuals' current functioning, and five risk-management items that account for professional opinions regarding the individuals' ability to adjust to either the institution (In ratings), or community (Out ratings). For the purpose of this study the "In ratings" were used as we were concerned with change in scores during their time in hospital. Each item is scored on a 3-point scale as either absent (0 or No), possibly present (1 or Possible), or definitely present (2 or Yes). The psychopathy item was omitted as the majority of patients had not been assessed with the PCL-R (15); this has been found to have minimal effect on the predictive efficacy of the HCR-20 (16) and the item has been removed from version 3 (14).

2.4.2. Patient characteristics

Information regarding the patients' age, gender, admission and discharge dates, legal status, security level, self-reported ethnicity, and ICD-10 (17) psychiatric diagnoses were extracted from patient records. In the UK secure mental health care is provided at three

levels: high security provides for patients assessed as posing a 'grave and immediate danger to the public'; medium security is suitable for patients who pose a 'serious danger to the public' while low secure is for those who pose a 'significant risk to others or themselves' (18). Security levels differ in the amount of physical and procedural security that is implemented to maintain safety. Civil commitments provide for assessment and/or treatment of those deemed at risk to themselves or others, while forensic commitments provide for assessment and/or treatment for patients diverted from criminal justice services or transferred temporarily from prison; informal patients are those admitted voluntarily and therefore not detained under the Mental Health Act (1983, Amended 2007). There were a large number of patients with both schizophrenia and personality disorder and they were therefore treated as a distinct comorbid group in the analysis.

2.4.3. Aggression data

Incidents of aggression in the three month period following the first HCR-20 assessment was extracted from patient records and coded using the Overt Aggression Scale (OAS; 19) as part of a previous study (20). Rating was conducted by three trained researchers blind to the HCR-20 risk assessments at the point of coding; inter-rater reliability was in the excellent range (Cohen's Kappa 0.81-1.00). For the purpose of the current paper, we were only interested in physical aggression against others. This was dichotomised as present/absent based on meeting or exceeding the criteria for a level 1 incident on the OAS (makes threatening gesture, swings at people, grabs at clothes).

2.5. Data analysis

Descriptive statistics were conducted to examine the prevalence of sample characteristics, HCR-20 summary judgment ratings, mean HCR-20 total, historical, clinical, and risk-management scores, and the presence of physical aggression towards others. Linear

mixed models, using maximum likelihood estimation, were used to examine baseline scores and rates of change in HCR-20 total, historical, clinical, risk-management, and individual item scores across assessments. Differences in rates of change between demographic and clinically defined groups were also examined. In order to examine whether time between assessments affected rates of change linear mixed models were also used to identify differences between individuals whose time between assessments fell outside of two standard deviations from the mean and those whose did not. Mixed effects models were used as they allow for consideration of change in variables over time, differences in mean scores between groups, and interactions between time and group (i.e., whether rates of change differ between groups). Further, mixed effects models account for varying number of assessments per patient thus maximising sample size; in contrast, repeated measures ANOVA models require removal of an entire patients' dataset (listwise deletion) if data is missing for any time point (21). Assessment iteration was entered as the repeated measures variable, and was coded such that estimated intercept values could be interpreted as baseline scores, corresponding to the first assessment. Fixed effects are used to model the variables of interest (13); therefore, gender (reference category = male), age (reference category = aged less than 40 years), ethnicity (reference category = Caucasian), diagnosis (reference category = schizophrenia), intellectual disability (reference category = no intellectual disability), substance abuse (reference category = no substance abuse), Mental Health Act status (reference category = civil), security level of ward (reference category = low secure), and presence or absence of physical aggression towards others in the three months following the first HCR-20 assessment (reference category = absent) were entered as fixed variables. Further, fixed effects were estimated for the interaction between assessment iteration and each of the factors; the resulting estimates correspond to the change in number of points with each iteration, relative to the reference group. The repeated HCR-20 assessments were modelled

with a random effect, to account for heterogeneity among patients' scores (22). Cohen's *d* values were calculated to examine the effect size of changes in scores and differences in baseline scores between groups; 0.2, 0.5, and 0.8 are typically considered the thresholds for small, medium, and large effect sizes.

Finally, the Reliable Change Index (RCI; 23) was calculated to determine the number of individuals whose scores have changed sufficiently to be confident that observed change is not due to measurement error. More specifically, reliable change is the extent to which the change measured for an individual is beyond that which could be due to the measurement variability of the instrument used. The measurement variability, or RCI, is determined by calculating the SE of the difference, with the following formula:

$$SE_{diff} = SD_1\sqrt{2\sqrt{1} - r}$$

SD_1 is the standard deviation of the baseline scores on the measure and r is the reliability of the measure. Change in scores that exceed 1.96 times the *SE* lie outside 95% of the area under a normal distribution curve and is thus unlikely to occur more than 5% of the time by unreliability of the measure alone. Internal consistency, as indicated by Cronbach's α derived from HCR-20 ratings in the current study, was used as the measure of reliability (historical scale = .521, clinical scale = .614, risk-management scale = .738. Clinically significant change, defined as a change in score from a dysfunctional population to a normal population (23), was also calculated. Given the lack of available HCR-20 scores in a normal population (8), scores were compared against HCR-20 assessments rated at the point of discharge from medium secure units in the UK (24). Therefore, our operational definition for a clinically significant change in this study was movement from scale scores above previously published data for patients deemed to no longer require inpatient care to scores equal to or below those cut-offs. Cut-off points were established for the HCR-20 total score (25.3), historical scale

(12.7), clinical scale (5.0), and risk-management scale (5.1) based on criteria C outlined by Evans, Margison, and Barkham (23); this approach is recommended when there is an overlap in distribution of scores between the two populations. The number of individuals whose summary judgement risk level changed was calculated as a second measure of change that may be considered clinically significant; the HCR-20 authors recommend (5) that these summary judgments are used in clinical practice when communicating risk, as opposed to the numerical ratings, since different items may have a differential weight depending on the case at hand. All statistics were conducted using PASW Statistics version 18 for Windows (25).

3. Results

3.1. Sample characteristics

There were 523 patients who had at least two HCR-20 assessments; 43 were excluded due to excessive missing information from assessments leaving a final sample size of 480. The mean time between the first and fourth assessments was 530.3 days (SD=61.8). Two thirds of the sample were male (*n*=321; 66.9%) and one third female (*n*=159, 33.1%). Mean age at assessment was 38.4 years (*SD*=15.6); for the purpose of analysis age was categorised as below 40 years of age (*n*=281, 58.5%) and aged 40 years and above (*n*=199, 41.5%).There were 225 Caucasian patients (46.9%), 35 Black patients (72.9%), 20 Asian patients (41.7%), 25 patient with a mixed ethnic background (5.2%), 2 patients from an “other ethnic background" (0.4%), and 173 patients (36.0%) with no recorded ethnicity. For the purpose of analysis, these were grouped into Caucasian (*n*=225, 46.9%), non-Caucasian (*n*=82, 17.1%), and unknown (*n*=173, 36.0%). The most common ICD-10 [16] psychiatric diagnoses were “Schizophrenia, schizotypal and delusional disorders” (F20-29; *n*=256, 53.3%), “Disorders of adult personality and behaviour” (F60-69; n=206, 42.9%),“Mental retardation” (F70-79; n=91, 19.0%), “Disorders of psychological development” (F80-89; n=72, 15.0%), and “Organic, including symptomatic, mental disorders” (F00-09; n=40, 8.3%). For the purpose

of analysis, diagnoses of "Mental retardation" (intellectual disability) and "Mental and behavioural disorders due to psychoactive substance use" (substance abuse) were entered as a separate variable, coded as present or absent, to minimise the level of complex co-morbidity in the sample. Participants were categorised according to the remaining ICD-10 major categories based on diagnosis within a single ICD-10 category; however, there were a large number of patients with comorbid schizophrenia and personality disorder, and these were treated as a distinct group. There were also a large number of patients with multiple or uncommon diagnoses that could not be grouped in any meaningful way into sufficiently large groups for analysis. Therefore, the final diagnostic groups were schizophrenia (*n*=180, 37.5%), personality disorder (*n*=93, 19.4%), schizophrenia & personality disorder (*n*=40, 8.3%), developmental (*n*=29, 6.0%), organic (*n*=22, 4.6%), and other or multiple (*n*=116, 24.2%). At the time of assessment, 336 (70.0%) patients were resident in low secure wards and 144 (30.0%) were resident in medium secure wards. There were 188 patients (39.2%) admitted under civil sections of the Mental Health Act (1983, amended 2007), 271 (56.5%) admitted under forensic sections, and 21 (4.4%) informal patients. In the three months following the first HCR-20 assessment, 197 (41.0%) had engaged in physical aggression against others.

3.2. HCR-20 assessments

There were 162 (33.8%) patients who had only two HCR-20 assessments completed, 218 (45.4%) had three assessments, and 100 (20.8%) had four assessments. The mean time between assessments was 181.5 days (SD=48.3). Deviation from the mean time between the assessments was defined as i) none, if the time between all of the individuals' assessments fell within two standard deviations of the mean, ii) less, if the time between any of the individuals' assessments was shorter than the mean minus two standard deviations, and iii) more, if the period between any of the individuals' assessments was longer than the mean

plus two standard deviations. Examination of change in scores over time using deviation from the mean as a factor found that this did not affect rates of change for clinical (F[477,2]=0.65, p=.521), risk-management (F[427,2]=1.87, p=.156), historical (F[356,2]=1.68, p=.187), or total scores (F[425,2]=1.74, p=.177); therefore, all individuals were retained for the subsequent analyses.

Across all assessments conducted at each time point, HCR-20 total scores ranged from 7 to 38 (Mean = 27.22; SD = 5.61). Historical, clinical and risk-management scale scores all spanned the total possible range; the mean historical score was 13.50, clinical score mean was 6.45 (SD = 2.36), and the mean risk-management score was 6.75 (SD = 2.47). The HCR-20 total score and dynamic scale scores decreased across subsequent assessments; however, the historical scale did not show this reduction and actual changes were very small. At all time points, moderate risk was the most common level assigned by the summary judgment. The proportion of individuals rated as low risk increased over time, and the proportion rated as high risk decreased; the largest reduction in high risk ratings and increase in low risk ratings was between the third and fourth assessments (see Table 1).

3.3. Baseline scores

3.3.1. Historical

Baseline scores were 2.26 points lower in individuals with a developmental diagnosis and 3.25 points lower in those with organic diagnoses relative to those with schizophrenia (p<.001; d=.37, and p<.001; d=.47). Scores were also 1.28 points lower in patients admitted informally, compared with those admitted under civil sections of the Mental Health Act (p=.040; d=.18) and 1.41 points lower in individuals aged 40 years or over compared with those aged under 40 years (p<.001; d=.49). In contrast, scores were 0.67 points higher in medium secure patients compared with low secure (p=.018; d=.21), 1.57 points higher in individuals with comorbid schizophrenia and personality disorder compared with patients

with a sole diagnosis of schizophrenia (p=.001; d=.29), and 1.45 points higher in individuals with a substance abuse diagnosis, compared with those without (p<.001; d=.37).

3.3.2. Clinical

Baseline scores were 1.74 points lower (p<.001; d=.50) in those with a personality disorder, 1.41 points lower (p=.001; d=.30) in those with a developmental diagnosis, and 0.73 points lower (p=.004; d=.25) in those with other or multiple diagnoses compared with those with schizophrenia. Scores were also 0.89 points lower (p<.001; d=.40) in forensic patients and 2.00 points lower (p<.001; d=.37) in informal patients compared with civil patients. Conversely, scores were 0.77 points higher in patients resident in medium secure wards compared with low secure wards (p<.001; d=.31), 0.68 points higher in women compared with men (p=.002; d=.26), , and 1.02 points higher in those who had been physically aggressive compared with those who had not (p<.001; d=.43).

3.3.3. Risk-management

Baseline scores were 0.71 points higher in those resident in medium secure wards compared with those in low secure wards (p=.002; d=.27), 0.49 points higher in women compared with men (p=.048; d=.17), 0.48 points higher in those with an unrecorded ethnicity compared with Caucasians (p=.035; d=.18), 0.53 points higher in those aged 40 years or over compared with those aged less than 40 years (p=.013; d=.22), and 1.06 points higher in individuals that had been aggressive relative to those who had not (p<.001; d=.41).

3.3.4. Total

Scores were 1.64 points lower in people with a personality disorder, 3.64 points lower in people with a developmental diagnosis, and 2.90 points lower in people with an organic diagnosis, compared with individuals with schizophrenia (p=.028; d=.19, p<.001; d=.31, and

p=.012; *d*=.22, respectively). Scores were also 4.23 points lower in individual who were admitted informally, compared with civil patients (*p*<.001; *d*=.32). Scores were 2.33 points higher in patients resident in medium secure wards compared with those in low secure wards (*p*<.001; *d*=.39), 1.44 points higher in women compared with men (*p*=.010; *d*=.23), 1.83 points higher in individuals with a substance abuse diagnosis compared to those without (*p*=.006; *d*=.24), and 1.94 points higher in individuals who had been physically aggressive compared with those who had not (*p*<.001; *d*=.33).

3.4. Change in scores over time

Mean changes in HCR-20 total scores, subscale scores, and individual item scores are presented in Table 2.

3.4.1. Historical

There was no significant reduction in historical scores across assessments (0.05 points, *p*=.556; *d*=.07). Individuals admitted informally reduced by an additional 0.36 points compared with those admitted under civil sections of the Mental Health Act (*p*=.037; *d*=.17). In contrast, individuals with an organic diagnosis increased by 0.53 points per assessment compared with those with schizophrenia (*p*=.001; *d*=.28).

3.4.2. Clinical

Across all participants, clinical scores reduced by 0.42 points per assessment (*p*<.001; *d*=.36). Individuals resident in medium secure wards decreased by an additional 0.19 points per assessment relative to those in low secure wards (*p*=.040; *d*=.19). In contrast, those with an organic diagnosis increased by 0.57 points per assessment, relative to those with schizophrenia (*p*=.017; *d*=.19), and those with a substance abuse diagnosis increased by 0.25 points relative to those without a substance abuse diagnosis (*p*=.030; *d*=.21); individuals who

had been aggressive towards others increased by 0.33 points compared with those who had not been aggressive (p<.001; d=.31).

3.4.3. Risk-management

Across participants, there was no significant reduction in risk-management scores over time (-0.06 points per assessment; p=.583; d=.06). Medium secure patients decreased by an additional 0.27 points compared with low secure patients (p=.003; d=.28) and informal admissions decreased by an additional 0.57 points relative to patients admitted under civil sections of the Mental Health Act (p=.015; d=.19). Scores increased by 0.64 points for individuals with an organic diagnosis compared with those with schizophrenia (p=.005; d=.23).

3.4.4. Total

HCR-20 total scores reduced by an average of 0.42 points per assessment (p=.048; d=.20). Scores decreased by an additional 0.48 points per assessment in patients resident on medium secure wards relative to those on low secure wards (p=.007; d=.26). In contrast, scores increased by 1.54 points in those with an organic diagnosis compared with individuals with schizophrenia (p=.001; d=.27) and by 0.45 points in those who had been physically aggressive relative to those who had not (p=.013; d=.24).

3.4.5. Individual items

None of the individual historical items decreased significantly over repeated assessments. However, H3 (relationship instability) ratings for those aged 40 years and over decreased by an additional 0.04 points per assessment relative to younger patients; p=.021; d=.26). Mean scores for individuals with a developmental diagnosis decreased by a further 0.08 on item H5 (substance use problems; p=.017; d=.31) and 0.12 on H9 (personality

disorder; p=.005; d=.34) compared with those with a diagnosis of schizophrenia. Mean scores for individuals with an organic diagnosis increased by 0.17 on item H1 (previous violence; p<.001; d=.68), 0.26 on item H6 (major mental illness; p<.001; d=.43), and 0.17 on H10 (prior supervision failure, p=.001; d=.41). Individuals admitted under forensic sections of the Mental Health Act decreased by an additional 0.05 points and informal admissions decreased by 0.12 points on item H8 (early maladjustment) relative to those admitted under civil sections of the Mental Health Act (p=.005; d=.32 and p=.017; d=.20, respectively). Individuals admitted informally decreased by 0.12 points on item H10 relative to those admitted under civil sections of the Mental Health Act (p=.020; d=.24).

With the exception of C5 (unresponsive to treatment), all of the clinical items reduced significantly across repeated assessments. However, mean scores of the medium secure group decreased by an additional 0.08 points on item C5 compared with those resident on low secure wards (p=.021; d=.21). Mean scores of individuals who had been physically aggressive towards others increased by 0.06 on item C1 (lack of insight; p=.006; d=.25), 0.06 on C3 (active symptoms of major mental illness; p=.047; d=.19), 0.07 for item C4 (impulsivity; p=.023; d=.21), and 0.09 on C5 (unresponsive to treatment; p=.008; d=.24) compared to individuals who had not been physically aggressive. Scores also increased on C5 for individuals admitted under forensic sections of the Mental Health Act relative to those admitted under civil sections (0.07 points; p=.028; d=.21), and for individuals with a substance abuse diagnosis compared with those without (0.09 points; p=.020; d=.23). Mean scores of individuals who held other or multiple diagnoses increased by 0.07 on item C2 (negative attitudes; p=.044; d=.20) and 0.08 on item C3 (p=.037; d=.21) relative to individuals with schizophrenia. Finally, mean scores of individuals who had co-morbid schizophrenia and personality disorder diagnoses increased by 0.14 on item C3 compared with those with a sole diagnosis of schizophrenia (p=.006; d=.29).

None of the risk-management items decreased significantly over time. However, scores for individuals resident in medium secure wards decreased by a further 0.08 on item R4 (p=.005; d=.27) and 0.06 points on item R5 (stress; p=.023; d=.21) per assessment relative to low secure patients. Scores for those with a diagnosis of intellectual disability decreased by 0.09 on item R1 (plans lack feasibility; p=.017; d=.24) and 0.06 points on R3 (lack of personal support; p=.043; d=.22) per assessment compared with those without an intellectual disability; the intellectual disability group increased by 0.06 points on item R5 relative to those without intellectual disability (p=.048; d=.18). Scores for people admitted informally decreased by an additional 0.14 points on item R3 and 0.16 points on item R4 relative to those admitted under civil sections of the Mental Health Act (p=.047; d=.15, and p=.042; d=.16). The organic group increased by 0.25 points for item R1 (p=.003; d=.24) and 0.15 points for R5 (p=.021; d=.18) compared with the schizophrenia group. For item R4, women increased by 0.10 points relative to men (p=.002 ; d=.28).

3.4.6. Reliable and clinically significant change

The numbers of individuals who showed reliable and clinically significant change are presented in Table 3. There were three individuals who showed reliable reduction in scores between the first and third assessment (one on the risk-management scale and two on the total score) who could not show clinically significant change, as defined in the current study, as their scores for the first assessment were already below the cut-off point. Rates of change increased across repeated assessments, such that proportions of individuals demonstrating reliable and clinically significant change were greater between the first and fourth assessments than between the first and second or first and third assessments. However, a maximum of 3% of individuals showed clinically significant reliable reductions in HCR-20 total scores over the study period. Across all comparisons, there were a greater number of individuals who demonstrated reliable and clinically significant reduction in scores on the

clinical (maximum 6%) and risk-management scales (maximum 5%) compared with the historical scale (maximum 1%).

Across each repeated assessment, a greater number of individuals decreased in risk level, as recorded by the summary judgment, than increased. A maximum of 0.9% of individuals decreased by two risk levels and 10.1% decreased by one risk level. Of those individuals who had decreased in risk level by the fourth assessment, only one had also showed reliable and clinically significant change in HCR-20 scores (see Table 4).

4. Discussion

The current study was the first to examine differences in rates of change in HCR-20 scores between demographically and clinically defined groups in a medium and low secure psychiatric setting. Across the sample as a whole, total clinical scale scores reduced significantly with repeated assessment, as did all individual clinical scale items with the exception of C5 (unresponsive to treatment). Neither of the other scales nor any other individual items reduced significantly. These results suggest that clinical functioning improves over the course of inpatient treatment, but that coping skills and the ability to adjust to the institutional setting do not. Further, they lend support for the clinical scale as a variable risk factor but cast some doubt regarding whether the risk-management scale meets the same criteria (3). The overall group findings also support the idea that the historical scale comprises relatively static risk factors (3); however, small but statistically significant differences in rates of change between groups suggest that these items may not be as static as previously thought.

Effect sizes for changes in scores were small. This may seem surprising, or even disappointing, since ratings were made in routine practice by clinical professionals who were also responsible for formulating individualized treatment and management plans to reduce risk based on the same assessment. However, they slightly exceeded change in clinical

scores documented by previous research conducted in high security settings (12, 25). Given that greater reductions in risk levels were observed between the third and fourth assessments, it is possible that greater change would have been observed over a longer study period. Assessments in the current study were conducted over an average of 17 months; therefore future studies may wish to examine change in scores over a minimum period of two years to investigate this possibility. Alternatively, it may be that the dynamic HCR-20 items are not easily modified, or that clinical teams are not accurately identifying when changes in risk factors occur. The fact that the number of individuals whose overall assessed risk level reduced exceeded the number of individuals who showed a modestly defined clinically reliable significant reduction in scores lends support for the former. Practitioners appear to be reducing their overall perceptions of risk in the absence of reduction in HCR-20 scores, suggesting that they are sensitive to change but are basing their decision, at least in part, on factors outside the HCR-20. Alternatively, clinicians may be incorrectly adjusting their risk levels in the absence of real change in risk, as opposed to the items being insensitive to change.

Baseline scores were broadly consistent with expectations such that medium secure patients and patients who were aggressive following the first HCR-20 assessment had higher scores; this likely reflects a need for increased security and greater risk of aggressive behaviour as the number of risk factors held by an individual increases. Conversely, scores were lower for informal admissions. This is consistent with evidence that the HCR-20 is able to differentiate between individuals at high risk of aggression, and those that are less likely to engage in such behaviour (8, 9, 20). Scores were also higher in individuals with schizophrenia, those with co-morbid schizophrenia and personality disorder, and those with substance abuse disorders. This is perhaps unsurprising given that three of the historical items

on the HCR-20 correspond to major mental illness, personality disorder, and substance abuse (5).

Items C1 (lack of insight), C2 (negative attitudes), C3 (active symptoms of major mental illness), and C4 (impulsivity) showed the greatest reduction over repeated assessments. It may be that these items are those that are most amenable to change. Alternatively, it may be that they are the items that are most actively being targeted by interventions given the clinical, as opposed to correctional setting.

Scores for individuals who were aggressive following the first HCR-20 assessment increased on all clinical items across repeated assessments, with the exception of C2 (negative attitudes). This is as expected and suggests that aggressive behaviour may have in part been due to an increase in symptoms and worsening clinical presentation, as evidenced by increased clinical scale scores. This study had a low threshold for the presence of physical aggression (level 1 on the OAS); therefore, it is possible that even larger increases in clinical scores may have been observed for individuals engaging in more serious aggression. Individuals with more complex diagnoses, such as comorbid personality disorder and schizophrenia, multiple diagnoses, and comorbid substance abuse also tended to increase in scores on clinical items over time.

Those with an organic diagnosis had lower HCR-20 total and historical scale scores at baseline compared with those with a diagnosis of schizophrenia; however, the organic group showed an increase in total and subscale scores, relative to the schizophrenia group. These findings suggest that the perceived risk of individuals with an organic diagnosis actually increases over the course of treatment This is perhaps not surprising as 63.6% of the patients with an organic diagnosis had a progressive neurological condition such as dementia or Huntington's disease. The greatest increase in scores for this group was observed for H6 (major mental illness), and R1 (plans lack feasibility); they also increased significantly on

items H1 (previous violence), H10 (prior supervision failure), and R5 (stress), relative to the schizophrenia group. However, the predictive validity of the HCR-20 for aggression among individuals with organic disorders has been found to be inferior to results achieved in individuals with personality disorders and schizophrenia (20). This suggests that the HCR-20 may not contain the most relevant risk factors for aggression among this population and, therefore, increases in scores may not necessarily result in increases in aggressive behaviour.

In contrast with previous findings (12, 26), women did not show a greater reduction in scores over time than men; they actually increased in scores on item R4 (noncompliance with remediation attempts) and had higher baseline total and dynamic scale scores compared with men. Reasons for these differences in findings are unclear. However, the majority of women (*n*=116; 73.0%) in the current sample had a personality disorder diagnosis (with or without other co-morbidities). Personality disorders are pervasive, persistent and stable (17) and individuals with such diagnoses may be less likely to show a reduction in risk factors over time. Neither Morriseey et al. (13) or Olsson et al. (12) present the diagnoses of their female sample. However, they may be expected to have demonstrated greater rates of change if their samples contain fewer individuals with personality disorder than the current study.

4.1. Limitations

The psychopathy item was excluded due to the vast majority of patients not being assessed with the PCL-R due to the additional time and expertise required. Whilst this has been found to have a negligible effect on predictive efficacy of the HCR-20 (16), this was not the focus of the current study. However, psychopathy has traditionally been viewed as an untreatable condition (27). Although this may be due to lack of effective treatments rather than people with the disorder being fundamentally untreatable (28), it suggests that this item may be unlikely to show much change over time. The study was naturalistic and pseudo-prospective and, as a result, HCR-20 ratings from routine clinical practice were used. It is

possible that different raters completed assessments over time and we were unable to ascertain inter-reliability ratings for different raters. It is also possible that raters may consciously or unconsciously rate patients at lower risk over time to reflect assumed treatment progress. Finally, it may be surprising that Risk Management items and total score did not change over time since these are intended to be dynamic factors and may be modifiable through intervention. We used 'In' (risk in the inpatient situation) rather than 'Out' (risk in the community) ratings and it is possible that this contributed to temporal stasis for these items since patients would continue to have high levels of support under this conjectured condition.

4.2. Implications and future considerations

The current study has extended previous findings about change in risk over time as measured with the HCR-20. Very few patients showed reliable clinical change over the study period. However, given the time period over which the current study was conducted, more longitudinal research is needed to see if change accelerates or is sustained following lengthy periods of treatment. Future research should also establish whether any change in scores is associated with a reduction in incidents. Summary judgements may have been slightly better indicators of reliable change than scores; however, where judgements fell scores did not which suggests that the HCR-20 scores are not sensitive to change and clinical raters are basing their overall assessment of risk on factors other than those contained in the HCR-20. Therefore, it would be inadvisable to make major changes to risk management, including changes in security level, based on changes in HCR-20 score; more focused assessments, such as the DUNDRUM toolkit (29), should be used for this purpose.

Table 1: Distribution of HCR-20 total, scale scores, and summary judgment ratings as a function of assessment

	Mean Scores (SD)				Summary judgment risk rating		
Assessment	Total	Historical	Clinical	Risk-management	Low	Moderate	High
1	27.49 (5.69)	13.50 (3.12)	6.81 (2.32)	6.81 (2.49)	80 (23.7%)	157 (46.4%)	101 (29.9%)
2	27.33 (5.56)	13.49 (3.11)	6.72 (2.29)	6.81 (2.43)	91 (25.9%)	158 (44.9%)	103 (29.3%)
3	26.97 (5.58)	13.53 (2.82)	6.43 (2.47)	6.71 (2.53)	64 (28.4%)	108 (48.0%)	53 (23.6%)
4	26.24 (5.52)	13.49 (2.95)	6.23 (2.40)	6.22 (2.39)	20 (40.0%)	23 (46.0%)	7 (14.0%)

Table 2: Mean change in scores per assessment

Item	Estimate	Std. Error	95%CI	*p*	*d*
H1	0.00	0.01	[-0.02, 0.02]	.980	.00
H2	0.03	0.02	[-0.01, 0.23]	.114	.19
H3	0.01	0.02	[-0.03, 0.05]	.690	.05
H4	-0.01	0.02	[-0.06, 0.03]	.519	.07
H5	0.02	0.02	[-0.02, 0.06]	.313	.10
H6	-0.01	0.02	[-0.06, 0.04]	.608	.05
H8	0.03	0.02	[-0.01, 0.08]	.137	.18
H9	-0.00	0.03	[-0.06, 0.05]	.920	.01
H10	-0.02	0.02	[-0.06, 0.02]	.330	.16
Historical total	0.05	0.08	[-0.11, 0.20]	.556	.07
C1	-0.08	0.03	[-0.13, -0.03]	.004	.28
C2	-0.10	0.04	[-0.17, -0.03]	.008	.27
C3	-0.10	0.04	[-0.17, -0.02]	.015	.25
C4	-0.10	0.04	[-0.16, -0.02]	.011	.25
C5	-0.06	0.04	[-0.14, 0.02]	.125	.15
Clinical total	-0.42	0.11	[-0.64, -0.20]	<.001	.36
R1	-0.06	0.04	[-0.14, 0.02]	.113	.16
R2	0.03	0.03	[-0.03, 0.10]	.289	.12
R3	0.01	0.03	[-0.05, 0.07]	.679	.05
R4	-0.05	0.04	[-0.11, 0.02]	.201	.13
R5	-0.00	0.03	[-0.06, 0.06]	.963	.00
Risk-management total	-0.06	0.11	[-0.27, 0.15]	.583	.06
HCR-20 total	-0.42	0.21	[-0.84, -0.01]	.048	.20

Table 3: Number of individuals showing reliable and clinically significant reductions in scores

	Assessment 1-2		Assessment 1-3		Assessment 1-4	
	RC (%)	R&CSC	RC (%)	R&CSC	RC (%)	R&CSC
Historical	2 (0.4%)	2 (0.4%)	1 (0.3%)	1 (0.3%)	1 (1.0%)	1 (1.0%)
Clinical	3 (0.6%)	3 (0.6%)	6 (1.9%)	6 (1.9%)	6 (6.0%)	6 (6.0%)
Risk-management	9 (1.9%)	8 (1.7%)	12 (3.8%)	10 (3.1%)	7 (7.0%)	5 (5.0%)
Total	6 (1.3%)	5 (1.2%)	6 (1.9%)	4 (1.3%)	4 (4.0%)	3 (3.0%)

RC, reliable change; R&CSC, reliable and clinically significant change

Table 4: Number of individuals whose summary judgment risk level reduced

	Assessment 1-2	Assessment 1-3	Assessment 1-4
Decreased			
1 level	25 (5.2%)	32 (10.1%)	8 (8.0%)
2 levels	1 (0.2%)	3 (0.9%)	0 (0.0%)
Increased			
1 level	8 (1.7%)	9 (2.8%)	2 (2.0%)
2 levels	2 (0.4%)	0 (0.0%)	0 (0.0%)

References

1. National Institute for Mental Health in England. Mental Health Policy Implementation Guide. Retrieved 1st March 2012 from http://wwwdhgovuk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4087136 [Internet]. 2004.

2. Webster CD, Nicholls TL, Martin ML, Desmarais SL, Brink J. Short-Term Assessment of Risk and Treatability (START): The case for a new structured professional judgment scheme. Behavioral Sciences and the Law. 2006;24(6):747-66.

3. Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, Kupfer DJ. Coming to terms with the terms of risk. Archives of General Psychiatry. 1997;54(4):337-43.

4. Douglas KS, Kropp PR. A prevention-based paradigm for violence risk assessment - Clinical and research applications. Criminal Justice and Behavior. 2002;29(5):617-58.

5. Webster CD, Douglas KS, Eaves D, Hart SD. HCR-20: Assessing risk for violence (Version 2). Burnaby, British Columbia: Mental Health, Law and Policy Institute, Simon Fraser University; 1997.

6. Khiroya R, Weaver T, Maden T. Use and perceived utility of structured violence risk assessments in English medium secure forensic units. Psychiatric Bulletin. 2009;33:129-32.

7. Douglas KS, Hart SD, Webster CD, Belfrage H, Guy LS, Wilson CM. Historical-Clinical-Risk Management-20, Version 3 (HCR-20^{V3}): Development and Overview. International Journal of Forensic Mental Health. 2014;13(2):93-108.

8. Douglas KS, Shaffer C, Blanchard AJE, Guy LS, Reeves KA, Weir J. HCR-20 violence risk assessment scheme: Overview and annotated bibliography (current up to January 1, 2014) 2014. Accessed 5 July 2015 at

https://kdouglas.files.wordpress.com/2014/01/hcr-20-annotated-bibliography-version-12-january-20142.pdf

9. O'Shea LE, Mitchell AE, Picchioni MM, Dickens GL. Moderators of the predictive efficacy of the historical, clinical and risk management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. Aggression and Violent Behavior. 2013;18:255-70.

10. Belfrage H, Douglas KS. Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. The International Journal of Forensic Mental Health. 2002;1(1):25-36.

11. Sheldon K, Gallagher J. S13-04 - Risk reductions in dangerous and severe personality disordered forensic in-patients measured with the HCR-20 violence risk assessment scheme. European Psychiatry. 2010;25, Supplement 1(0):35.

12. Olsson H, Strand S, Kristiansen L, Sjöling M, Asplund K. Decreased Risk for Violence in Patients Admitted to Forensic Care, Measured With the HCR-20. Archives of Psychiatric Nursing. 2013;27(4):191-7.

13. Morrissey C, Beeley C, Milton J. Longitudinal HCR-20 scores in a high-secure psychiatric hospital. Criminal Behaviour and Mental Health, 2015;24(3):169–80. doi:10.1002/cbm.1893.

14. Douglas K, Hart S, Webster C, Belfrage H. HCR-20: Assessing risk for violence (Version 3). Burnaby, BC: Mental Health, Law, and Policy Institute, Simon Fraser University. 2013.

15. Hare RD. Manual for the Hare Psychopathy Checklist - Revised. Toronto: Multi-Health Systems, Inc; 1991.

16. Guy LS, Douglas KS, Hendry MC. The role of psychopathic personality disorder in violence risk assessments using the HCR-20. Journal of Personality Disorders. 2010;24(5):551-80.

17. World Health Organisation. The ICD-10 classification of mental and behavioural disorders. Geneva: WHO; 1992.

18. Centre for Mental Health. Secure care: a briefing note. London: Centre for Mental Health; 2013. Accessed 5 July 2015 file:///C:/Users/Geoff/Downloads/briefing_note_secure_care.pdf

19. Yudofsky SC, Silver JM, Jackson W, Endicott J, Williams D. The Overt Aggression Scale for the objective rating of verbal and physical aggression. The American Journal of Psychiatry. 1986;143(1):35-9.

20. O'Shea LE, Picchioni MM, Mason FL, Sugarman PA, Dickens GL. Differential predictive validity of HCR-20 for inpatient aggression. Psychiatry Research. 2014;220:669-78.

21. West BT, Welch KB, Galecki AT. Linear mixed models: a practical guide using statistical software. London: CRC Press; 2006.

22. Antonio K, Beirlant J. Actuarial statistics with generalized linear mixed models. Insurance: Mathematics and Economics. 2007;40(1):58-76.

23. Evans C, Margison F, Barkham M. The contribution of reliable and clinically significant change methods to evidence-based mental health. Evidence-Based Mental Health. 1998;1:70-72.

24. Dolan M, Khawaja A. The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. Aggressive Behavior. 2004;30(6):469-83.

25. SPSS Inc. PASW Statistics for Windows (Version 18). Chicago: SPSS Inc; 2009.

26. Morrissey C, Hogue T, Mooney P, Allen C, Johnston S, Hollin C, et al. Predictive validity of the PCL-R in offenders with intellectual disability in a high secure hospital setting: Institutional aggression. Journal of Forensic Psychiatry & Psychology. 2007;18(1):1-15.

27. Harris GT, Rice ME. Treatment of psychopathy: A review of empirical findings. In: Patrick CJ, editor. Handbook of Psychopathy. New York: Guilford Press; 2006. p. 555-72.

28. Moreira D, Almeida F, Pinto M, Fávero M. Psychopathy: A comprehensive review of its assessment and intervention. Aggression and Violent Behavior. 2014;19(3):191-5.

29. Kennedy H, O'Neill C, Flynn G, Gill P. The Dundrum Toolkit. Dangerousness, Understanding, Recovery and Urgency Manual (The Dundrum Quartet) V1 021 Four Structured Professional Judgment Instruments for Admission Triage, Urgency, Treatment Completion and Recovery Assessments. Dublin, Ireland: Trinity College Dublin;2010.